Emergence of Modularity in Genotype-Phenotype Mappings

Vladimír Kvasnička
Jiří Pospíchal
Department of Mathematics
Slovak Technical University
81237 Bratislava, Slovakia
kvasnic@cvt.stuba.sk
pospich@cvt.stuba.sk

Abstract A novel evolutionary method that allows us to study the emergence of modularity for genotype-phenotype mapping in the course of Darwinian evolution is described. The method is based on composite epigenotypes with two parts: a binary genotype; and a mapping of genes onto phenotype characters. For such generalized epigenotypes the modularity is determined in the following intuitive way: The genes are divided into two subgroups; simultaneously with this decomposition there is defined an accompanying decomposition of the set of phenotype characters. We expect that for epigenotypes with modular structures the genes from one group will be mapped onto characters from the same group, that is, that the appearance of crosslink mappings will be maximally suppressed. A fundamental question for all of evolutionary biology (and also for evolutionary algorithms and connectionist cognitive science) is the mechanism of evolutionary emergence of modular structures. The presented explanatory model is an implementation of the assumption that variation in genotype is produced on a faster time scale than variation in the genotype-phenotype mapped part. Moreover, the evaluation of the epigenotype in the evolutionary algorithm is based on directly selectable properties (corresponding to the decomposition of the set of phenotype characters). The modularity of genotype-phenotype mapping emerges in the simulations.

Keywords
Modularity, genotype-phenotype mapping, evolutionary simulation, evolutionary algorithms

1 Introduction

Recently, much research has focused on the existence of developmental and evolutionary modules that are necessary for the evolution of morphological phenotypes [1–3, 8, 9, 11, 23–25]. The emergence of modules is considered a very important part of the process that led to the evolutionary formation of complex organisms. Unfortunately, the main reasons for the emergence of modularity and its evolutionary mechanisms are very little understood; the few theories published [6, 7, 10, 26] about the origin of modules point in very different directions.

The purpose of this paper is to design a simple model of evolutionary emergence of modularity for genotype-phenotype mapping. Simple binary-string chromosomes [16, 17] are generalized in this model to a formal structure composed of (1) a genotype part represented by a binary string of fixed length and (2) a part represented by a mapping of simple genes onto phenotype characters (the so-called genotype-phenotype mapping) [23–26]. These structures, here called *epigenotypes*, are of the minimal complexity sufficient to allow a meaningful definition of modularity. By modularity we intuitively understand a division of genes and characters into two disjoint and nonempty sub-

sets such that the occurrence of *crosslink* mappings of genes from one subset (say A) onto another subset (say B) of phenotype characters is minimal [ideally, genes from $A\,(B)$ are mapped only onto characters from $A\,(B)$]. Our main interest is in formulating an evolutionary algorithm where modularity will emerge without an explicit built-in specification of evolutionary advantages of epigenotypes with modular structures. In particular, the Darwinian evolution used (represented numerically by a simple version of evolutionary algorithms) employs natural selection based on the so-called effective fitness, which reflects to some extent a quality-of-fitness landscape imposed by the phenotype mapping, which influences the adaptability of the genotype part. Adaptability in this context means ability to achieve a better solution by mutation of the original genotype part.

Recent theoretical approaches to modularity [6, 7] suggest its origin in a side effect of environment specialization. Numerical results presented in this paper give rise to the firm conclusion that the emergence of modularity is a by-product of some sort of congruence between modularity and directly selectable properties [26]. Here we have observed congruence between modularity of genotype-phenotype mapping and the variation rate of genotypes.

2 General Theory of Modularity

Binary vectors of fixed length n represent a *genotype*

$$G = (g_1, g_2, \ldots, g_n) \in \{0, 1\}^n \tag{1}$$

whose components are called *genes*. A *phenotype* P is represented by a set of indices on the dimensions or characters of the phenotype. This set is composed of the first m positive integers,

$$P = \{1, 2, \ldots, m\}. \tag{2}$$

This means that our hypothetical agent has a phenotype composed of m *characters* (*traits*).

One of fundamental concepts of our approach is the so-called *genotype-phenotype index mapping*

$$\Gamma\colon\ \{1, 2, \ldots, n\} \to 2^{\{1,2,\ldots,m\}}. \tag{3}$$

In this mapping, each gene index $i \in \{1, 2, \ldots, n\}$ is evaluated by a subset of character indices,

$$\Gamma(i) = \{j_1 < j_2 < \cdots < j_{r_i}\}.$$

An *inverse mapping* to this genotype-phenotype index mapping is determined by

$$\Gamma^{-1}\colon\ \{1, 2, \ldots, m\} \to 2^{\{1,2,\ldots,n\}}, \tag{4}$$

where each character index $j \in \{1, 2, \ldots, m\}$ is evaluated by a subset of gene indices $\Gamma^{-1}(j) = \{i_1 < i_2 < \cdots < i_{s_j}\}$. A subset $\Gamma^{-1}(j)$ is interpreted as saying that a phenotype character j is specified by genes $i_1, i_2, \ldots, i_{s_j}$. We say that the Γ and Γ^{-1} are mutually *inverse*; formally these mappings simultaneously satisfy

$$\forall j \in \{1, 2, \ldots, m\} \quad \forall i \in \Gamma^{-1}(j)\colon \qquad j \in \Gamma(i). \tag{5}$$

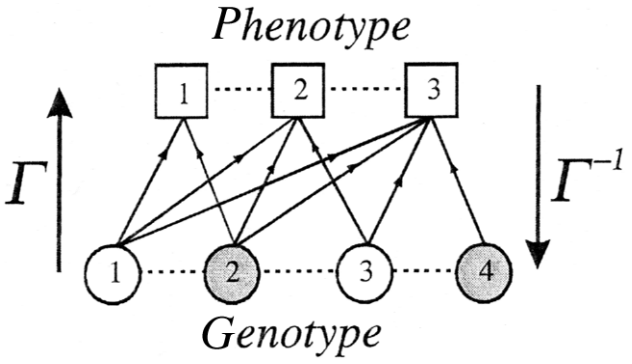

Figure 1. An illustrative example of genotype-phenotype index mapping of a genotype composed of four components (genes) onto a phenotype composed of three characters. Dark (light) genes correspond to binary allelic values 1 (0).

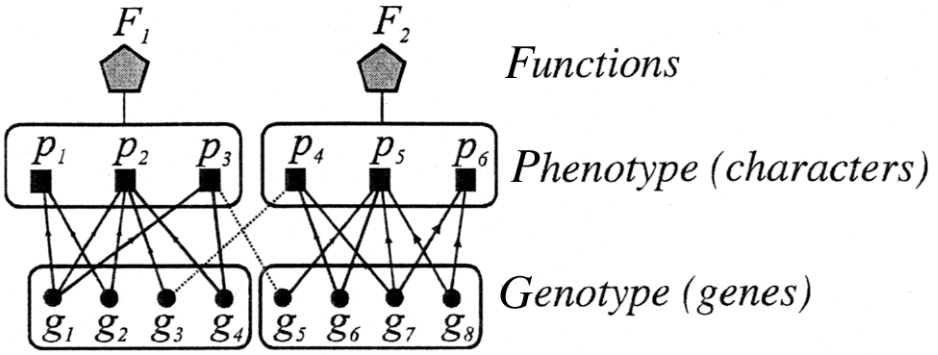

Figure 2. An example of a modular representation of the character complexes (functions) $F_1 = \{p_1, p_2, p_3\}$ and $F_2 = \{p_4, p_5, p_6\}$. The left (right) gene group only weakly influences F_2 (F_1). The genetic representation is modular in that the left-hand genes $\{g_1, g_2, g_3\}$ primarily affect the left-hand side characters, and the right-hand genes $\{g_4, g_5, g_6\}$ the right-hand side characters. There are more effects on the characters within each complex (represented by solid lines) than between them (dotted lines).

An illustrative example is presented in Figure 1; the corresponding sets Γ and Γ^{-1} are given by $\Gamma(1) = \{1, 2, 3\}, \Gamma(2) = \{1, 2, 3\}, \Gamma(3) = \{2, 3\}, \Gamma(4) = \{3\}$, and by $\Gamma^{-1}(1) = \{1, 2\}, \Gamma^{-1}(2) = \{1, 2, 3\}, \Gamma^{-1}(3) = \{1, 2, 3, 4\}$.

The notion of *modularity* is defined in this context with respect to a particular genotype-phenotype index mapping. Loosely speaking, a mapping Γ is modular if its genotype and phenotype parts can be decomposed into smaller parts such that a given genotype subpart (gene module) can be mapped almost entirely onto a phenotype subpart (phenotype module), whereas crosslink mappings are kept to a minimum (see Figure 2).

A more exact version of the above intuitive definition of modularity is as follows (in the graph theory literature such a problem, similar to our notion of modularity, is called the graph decomposition problem [19]): Let us consider a bipartite graph $\Gamma = (V = G \cup P, \ E \subseteq G \times P)$ composed of a vertex set V identified with genes and characters; its edge set E is composed of oriented edges pointing from genes to characters (cf. Figures 1 and 2). Let us assume that sets G and P are decomposed onto two disjoint subsets

$$G = G_A \cup G_B \quad \text{and} \quad P = P_A \cup P_B. \tag{6}$$

For these decompositions we define the following edge subsets:

$$E_{XY} = \{(u, v) \in E; \quad u \in G_X \quad \wedge \quad v \in P_Y\} \qquad (X, Y = A, B). \tag{7}$$

For instance, a subset $E_{AA}\,(E_{AB})$ contains all edges (crosslinks) that start at the gene subset G_A and terminate at the phenotype subset $P_A\,(P_B)$. An actual form of the decompositions (6) is subject to the requirement

$$|E_{AB}| + |E_{BA}| = \min \tag{8a}$$

simultaneously with

$$\left(\frac{n}{2} - |G_A|\right)^2 = \min, \qquad \left(\frac{m}{2} - |P_A|\right)^2 = \min. \tag{8b}$$

The first condition (8a) corresponds to our requirement that the number of crosslink mappings between different modules be minimized (i.e., suppressed). The next two conditions (Equation 8b) manifest our intuitive requirement that the decompositions in Equation 6 divide each of the sets P and G into two subsets each of which is composed of roughly half the elements of the original set.

An *epigenotype*[1] assigned to a given pair of genotype $G = (g_1, g_2, \ldots, g_n)$ and mapping Γ is determined as the following composition (see Figure 3):

$$\boldsymbol{x} = ((g_1, \Gamma(1)), (g_2, \Gamma(2)), \ldots, (g_n, \Gamma(n))). \tag{9}$$

In standard evolutionary algorithms, chromosomes (here replaced by epigenotypes) do not contain a mapping of genotype onto phenotype, or in other words, we may say that the mapping is kept fixed over the whole course of the evolution.

An *epigenotype domain* $X = \{\boldsymbol{x}\}$ composed of all possible epigenotypes is determined as the Cartesian product of the set of all binary vectors of length n and the set composed of all mappings Γ:

$$X = \underbrace{\{0, 1\}^n}_{\substack{\text{genotype}\\ \text{part}}} \times \underbrace{(2^{\{1,2,\ldots,m\}})^n}_{\substack{\text{phenotype}\\ \text{part}}}. \tag{10}$$

The cardinality of X is simply determined by

$$|X| = 2^n \times (2^m)^n = 2^{n(1+m)}. \tag{11}$$

Let f be the so-called *fitness function*, which evaluates each epigenotype $\boldsymbol{x} \in X$ by a positive real number $f(\boldsymbol{x})$, $f\colon X \to R_+$. Following S. Wright [27], Darwinian evolution may be interpreted as an optimization problem over a fitness landscape represented by the fitness function f:

$$\boldsymbol{x}_{\text{opt}} = \arg\max_{\boldsymbol{x} \in X} f(\boldsymbol{x}), \tag{12}$$

where the required number of points, N_{search}, needing to be sampled under exhaustive search through the genetic algorithm (GA) domain to get the solution is determined

1 The epigenotype corresponds to a unit of information, which is subject to adaptation like that of a chromosome in genetic algorithms [16, 17]. This term was suggested by one of the referees.

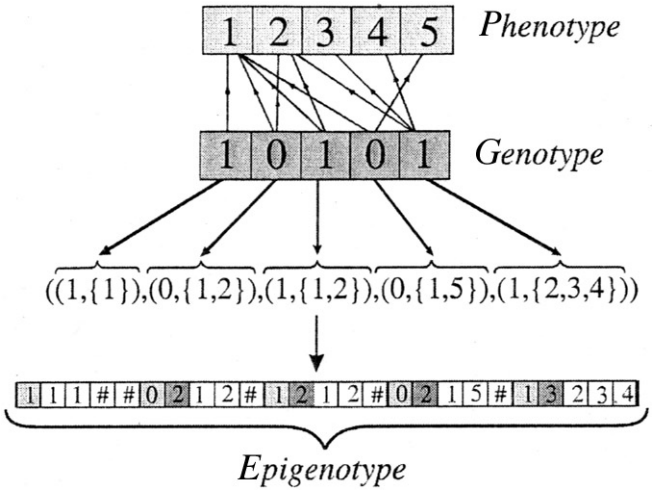

Figure 3. An epigenotype x is determined as a composition of the respective genotype (represented by a binary vector G) and the mapping Γ. The third row represents a formal representation of the epigenotype $x = (G, \Gamma)$ such that each gene g_i is accompanied by the respective mapping $\Gamma(i)$. The fourth row corresponds to a numerical representation of the epigenotype x; it is a linear string of nonnegative integers, and its single blocks are a numerical representation of $(g_i, \Gamma(i))$ for $i = 1, 2, \ldots, n$, where each gene is mapped here onto at most three characters.

by the cardinality of X (see Equation 11), $N_{\text{search}} \approx 2^{n(1+m)}$. The classical GA has the search set $X = \{0, 1\}^n$, which means that the total search through the GA domain requires $N_{\text{search}} \approx 2^n$. This result implies that a present generalization of the GA (using epigenotypes specified as a composition of binary string and mapping) is substantially more time consuming than its classical version.

The above fitness function will be specified as an analogue of *Kauffman's* NK *function* [4, 20]. An epigenotype $\boldsymbol{x} = ((g_1, \Gamma(1)), \ldots, (g_n, \Gamma(n)))$ is evaluated by the fitness $f(\boldsymbol{x})$ as follows:

$$f(\boldsymbol{x}) = \sum_{j=1}^{m} \varphi_{s_j}^{(j)}(\langle G, \Gamma^{-1}(j)\rangle), \tag{13}$$

where the argument $\langle G, \Gamma^{-1}(j)\rangle$ is a binary subvector composed of those entries of G that are elements of the inverse mapping $\Gamma^{-1}(j) = \{i_1, i_2, \ldots, i_{s_j}\}$. (The explicit dependence of $i_1, i_2, \ldots$ on j is not shown in this formalism, in order to abbreviate complicated expressions.) The variable s_j is the *polygeny* (the number of genes determining the phenotype character, i.e., the *indegree* of phenotype vertices) of the fitness component $\varphi^{(j)}$, and is derived from $\Gamma^{-1}(j)$:

$$\langle G, \Gamma^{-1}(j)\rangle = (g_{i_1}, g_{i_2}, \ldots, g_{i_{s_j}}). \tag{14a}$$

Then

$$f(\boldsymbol{x}) = \sum_{j=1}^{m} \varphi_{s_j}^{(j)}(g_{i_1}, g_{i_2}, \ldots, g_{i_{s_j}}). \tag{14b}$$

Details of the Kauffman's NK function used in our simulations are given in the Appendix.

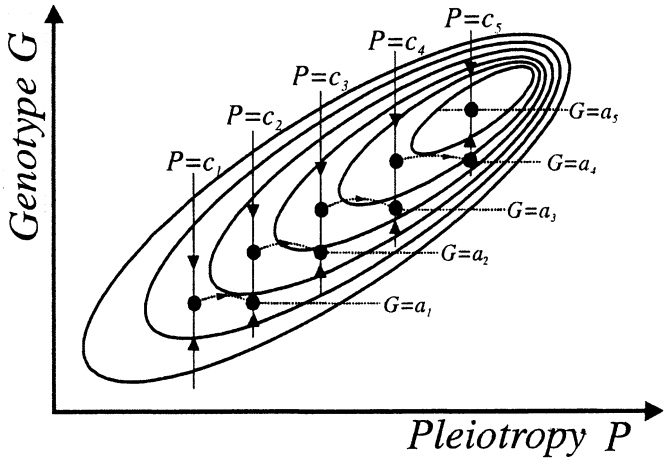

Figure 4. A schematic outline of an evolutionary process over a two-dimensional fitness landscape, where the first (second) variable corresponds to a pleiotropy (genotype). Since we have postulated in our specification of GA (Darwinian evolution in silico) that mutations of genes are substantially more frequent than mutations of genotype-phenotype mappings, a dominant part of evolution is running over a genotype adaptation based on gene variations (vertical), and only a very small part of evolution covers variations of the pleiotropy (horizontal). Loosely speaking, our Darwinian evolution of epigenotypes is composed of repeated sequences of long-term gene adaptation (where genes are adapted [optimized] for a fixed mapping) and rarely appearing pleiotropy adaptation. On the plot of the fitness landscape this mode of adaptation is represented by *adaptation stairs* toward the maximum. The ovals represent a *contour map* of a fitness function, with optimum fitness in the right upper corner. The arrows show a course of optimization, as the position of best-found epigenotype moves through the search space, where the c_i (a_i) are pleiotropy (genotype) locally optimal positions.

3 Short Time Scale Evolution of the Genotype

Let us assume that a goal of our Darwinian evolution is to solve the optimization problem in Equation 12 so that a particular modularity of final solution is achieved. This modularity would be in agreement with a given separation of phenotype characters into two subgroups. Since this problem is of extreme complexity (see Equation 13), we have to turn our attention to heuristics, which substantially simplify the process of finding an appropriate solution. We postulate that Darwinian evolution [23–26] of our epigenotypes is based on the biologically plausible assumption that genotype mutations are substantially more frequent than pleiotropy mutations (loosely speaking, mutations of a phenotype mapping) (see Figure 4):

1. An epigenotype is evolved so that its genotype composition is adapted while the respective mapping is kept fixed (it represents a constraint on genotype adaptation).
2. The mapping parts of epigenotypes are mutated very infrequently, while the respective gene compositions are *locally optimized* (adapted).

Why do we expect that in the course of Darwinian evolution of a population composed of our type of epigenotypes there will emerge ones with a modular structure? For epigenotypes with modular structure there might exist a selection pressure that gives rise to their spontaneous emergence. Of course, this fundamental requirement may be implemented in an explicit way in our simulation calculations; but we prefer to follow more intrinsic ways based on our assumption that modular structures indirectly support an increase of fitness through various mechanisms. For instance, an appearance

of modular structure in epigenotypes allows local optimizations within each module independent of other modules, which substantially accelerates the Darwinian evolution. Our first argument for the creation of modular structures is the likelihood of their creation by more or less stochastic occurrences. For a particular epigenotype $\boldsymbol{x}$ there exists a *parcellation* [24] such that if we remove two elementary mappings we get a new epigenotype x' with a higher fitness.

Let us consider different epigenotypes $x_1 = (\alpha_1, \Gamma_1)$ and $x_2 = (\alpha_2, \Gamma_2)$ with the same fitness, $f(x_1) = f(x_2)$. Let $U(x_1) = \{x' = O_{\text{mut}}(x_1)\}$ and $U(x_2) = \{x'' = O_{\text{mut}}(x_2)\}$ be neighborhoods of x_1 and x_2; they are composed of those epigenotypes that are mutations of the original epigenotypes. If in $U(x_1)$ an epigenotype x_1' exists such that $f(x_1') > f(x_1)$ and in $U(x_2)$ such an epigenotype does not exist, then we say that these epigenotypes x_1 and x_2 are distinguished from the standpoint of their possible genotype variation evolvability.

Let us assume that fitness $f(x)$ is specified as follows:

$$f(\boldsymbol{x}) = \cdots + \varphi_4^{(3)}(g_1, g_2, g_3, g_4) + \varphi_4^{(4)}(g_3, g_4, g_5, g_6) + \cdots . \tag{15a}$$

If the inequality

$$\varphi_4^{(3)}(g_1, g_2, g_3, g_4) + \varphi_4^{(4)}(g_3, g_4, g_5, g_6) \leq \varphi_3^{(3)}(g_1, g_2, g_3) + \varphi_3^{(4)}(g_4, g_5, g_6) \tag{15b}$$

is satisfied, then we may say that for the epigenotype x there exists a parcellation such that we get a new epigenotype x' with a modular structure and with a higher fitness, $f(\boldsymbol{x}) < f(\boldsymbol{x}')$. A partial probabilistic analysis of the probability that Equation 15b will be fulfilled under the mutation operator [16], which might identify critical elements in the dynamics, will be presented in a forthcoming publication.

In order to simplify our forthcoming considerations, let us postulate that the phenotype index set $P = \{1, 2, \ldots, m\}$ is decomposed into two disjoint subsets:

$$P = P_A \cup P_B = \{1, 2, \ldots, m_A\} \cup \{m_A + 1, \ldots, m\} \tag{16}$$

with $m = m_A + m_B$. Let us introduce two independent parts of the fitness function in Equation 14 with respect to this decomposition:

$$f_A(\boldsymbol{x}) = \sum_{j=1}^{m_A} \varphi_{s_j}^{(j)}(\langle G, \Gamma^{-1}(j)\rangle), \tag{17a}$$

$$f_B(\boldsymbol{x}) = \sum_{j=m_A+1}^{m} \varphi_{s_j}^{(j)}(\langle G, \Gamma^{-1}(j)\rangle). \tag{17b}$$

A neighborhood $U(\boldsymbol{x})$ of an epigenotype $\boldsymbol{x} = (G, \Gamma)$ is constructed in such a way that the pleiotropy part is kept fixed and only the genotype part is varied (mutated):

$$U(\boldsymbol{x}) = U(G, \Gamma) = \{(G', \Gamma);\ G' = O_{\text{mut}}(G)\}, \tag{18}$$

where $G' = O_{\text{mut}}(G)$ is a binary string constructed from the original one $\boldsymbol{\alpha}$ by flipping a few of its entries randomly to complementary values. In this neighborhood we look for a new solution $\boldsymbol{x}_{\text{opt}} = (G_{\text{opt}}, \Gamma) \in U(\boldsymbol{x})$ that provides simultaneously a better solution

for f_A that is no worse for f_B (and vice versa), that is,

$$f_A(\boldsymbol{x}_{\text{opt}}) > f_A(\boldsymbol{x}) \quad \text{and} \quad f_B(\boldsymbol{x}_{\text{opt}}) \geq f_B(\boldsymbol{x}) \tag{19a}$$

or

$$f_A(\boldsymbol{x}_{\text{opt}}) \geq f_A(\boldsymbol{x}) \quad \text{and} \quad f_B(\boldsymbol{x}_{\text{opt}}) > f_B(\boldsymbol{x}). \tag{19b}$$

This approach is similar to a hill-climbing algorithm for the adaptation of the genotype part given the pleiotropy part. Equations 18–20 are all concerned with allelic evolution, that is, evolution by allelic substitution at the genetic loci. We see that the constructed solution is the so-called Pareto optimal one [15]: both fitness subfunctions are no worse than the original solution $\boldsymbol{x}$. Of course, if such a solution does not exist, then we put $\boldsymbol{x}_{\text{opt}} = \boldsymbol{x}$. The optimal solution $\boldsymbol{x}_{\text{opt}}$ is used for an evaluation of epigenotypes, where the split of the evaluation into two functions, with arguments given by separation of phenotype characters into two subgroups, provides a basis for future emergence of modularity:

$$f(\boldsymbol{x}_{\text{opt}}) = f_A(\boldsymbol{x}_{\text{opt}}) + f_B(\boldsymbol{x}_{\text{opt}}). \tag{20}$$

The difference between the evaluations of the optimal solution $\boldsymbol{x}_{\text{opt}}$ and the original solution $\boldsymbol{x}$ reflects the degree of local genotype evolvability of the particular epigenotype; that is, if $f_A(\boldsymbol{x}_{\text{opt}}) + f_B(\boldsymbol{x}_{\text{opt}}) > f_A(\boldsymbol{x}) + f_B(\boldsymbol{x})$, then we may say that the epigenotype $\boldsymbol{x}$ is appropriately (with respect to a proposed decomposition in Equation 16) evolvable. Moreover, the resulting optimal epigenotype $\boldsymbol{x}_{\text{opt}}$ *replaces* the original epigenotype $\boldsymbol{x}\,(\boldsymbol{x} \leftarrow \boldsymbol{x}_{\text{opt}})$ in a population P specified as a multiset of epigenotypes. A simultaneous process of evaluation of epigenotypes in a population by the locally optimized fitness (Equation 20) together with a creation of a new population from optimal solutions is outlined in the following algorithm.

ALGORITHM 1:

```
Q:=∅;
for each x ∈ P do
begin construct neighborhood U(x);
      find in U(x) a Pareto optimal solution x_opt;
      f(x_opt):=f_A(x_opt)+f_B(x_opt);
      Q:=Q ∪ {x_opt};
end;
P:=Q;
```

In this algorithm the original population P is updated by a new population Q composed of new epigenotypes $\boldsymbol{x}_{\text{opt}}$ that are Pareto optimal with respect to the original epigenotypes $\boldsymbol{x}$ (the mapping part of epigenotypes in the course of local optimization is kept fixed). The cardinality of the neighborhood $U(\boldsymbol{x})$ is among the important constants of the method; in our simulation we kept it equal to a few hundred.

4 Long Time Scale Evolution of the Genotype-Phenotype Map

Darwinian evolution of a population of epigenotypes is simulated by a simple version of an evolutionary algorithm handling a variation of the pleiotropy part combined with a short time scale evolution of genotype parts of all epigenotypes toward their

optimal forms. This approach follows from the fact that the representation used for epigenotypes has a genotype part, which may be subject to local adaptation. The basic principles of the evolutionary algorithm applied to a population composed of epigenotypes are outlined in the following algorithm [16, 17]:

ALGORITHM 2:

```
time:=0;
P:=randomly generated population of epigenotypes;
while time<time_max do
begin time:=time+1;
      locally optimize genotype parts of all epigenotypes
      using Algorithm 1;
      Q:=∅;
      while |Q|<|P| do
      begin x_1:=O_select(P); x_2:=O_select(P);
            (x'_1,x'_2):=O_cross(x_1,x_2);
            x''_1:=O_mut(x'_1);x''_2:=O_mut(x'_2);
            Q:=Q ∪ {x''_1,x''_2};
      end;
      P:=Q;
end;
```

The algorithm is initiated on a randomly generated population composed of epigenotypes. An outer while loop with counter variable `time` is repeated $\texttt{time}_{\max}$ times. Within this outer loop all genotype parts of epigenotypes are replaced by their locally optimized versions (see Algorithm 1). An inner while loop is controlled by whether the cardinality of the offspring population is equal to the cardinality of the original parental population. The core of the reproduction process is composed of a quasirandom selection (performed by Goldberg's roulette wheel [16]) of two epigenotypes based on their fitness (see Equation 20): crossovers and mutations. The resulting epigenotypes form the offspring population. After termination of the inner while loop the parental population is updated by the offspring population.

We have to emphasize that a modification of epigenotypes to other epigenotypes (by mutations and crossovers) is divided into two different stages. In the first stage only genotype parts of epigenotypes are changed, while their mapping parts are kept fixed. On the other hand, in the second stage, where a usual evolutionary algorithm (similar to GA) is running, only mapping parts of epigenotypes are changed by mutations and crossovers (genotype parts are kept fixed).

The mutation and crossover operations of the pleiotropy part, like the mutations and crossovers of a canonical genetic algorithm, can decrease the fitness of epigenotypes (and most probably do most of the time). The fitness of the new epigenotypes is not evaluated immediately, but in a new cycle during the local optimization of genotype parts of the new epigenotypes.

- The *mutation operator* assigns stochastically to an original epigenotype $\boldsymbol{x}$ a new epigenotype $\boldsymbol{x}'$ such that the genotype parts are kept fixed:

$$\boldsymbol{x}' = O_{\mathrm{mut}}(\boldsymbol{x}), \tag{21a}$$

where

$$\boldsymbol{x} = ((g_1, \Gamma(1)), \ldots, (g_n, \Gamma(n))), \tag{21b}$$

$$\boldsymbol{x}' = ((g_1, \Gamma'(1)), \ldots, (g_n, \Gamma'(n))). \tag{21c}$$

Three different mutations $\Gamma(i) \to \Gamma'(i) = O_{\text{mut}}(\Gamma(i))$ are considered: (1) a randomly selected index is deleted, (2) a randomly generated index is added, and (3) a randomly selected index is changed to another randomly generated index. Each of these three distinct processes is performed with the same probability.

- The *crossover operator* assigns to a pair of quasirandomly selected parental epigenotypes (the probability of their selection is proportional to their fitness) another pair of offspring epigenotypes,

$$(\boldsymbol{x}', \boldsymbol{y}') = O_{\text{cross}}(\boldsymbol{x}, \boldsymbol{y}), \tag{22a}$$

where the parental epigenotypes are

$$\boldsymbol{x} = ((g_1, \Gamma(1)), \ldots, (g_i, \Gamma(i))(g_{i+1}, \Gamma(i+1)), \ldots, (g_n, \Gamma(n))), \tag{22b}$$

$$\boldsymbol{y} = ((g'_1, \Gamma'(1)), \ldots, (g'_i, \Gamma'(i))(g'_{i+1}, \Gamma'(i+1)), \ldots, (g'_n, \Gamma'(n))), \tag{22c}$$

and new offspring epigenotypes are created from the old ones in such a way that beyond a crosspoint index i (randomly selected) the remaining parts of parental mappings are exchanged:

$$\boldsymbol{x}' = ((g_1, \Gamma(1)), \ldots, (g_i, \Gamma(i)), (g_{i+1}, \Gamma'(i+1)), \ldots, (g_n, \Gamma'(n))), \tag{22d}$$

$$\boldsymbol{y}' = ((g'_1, \Gamma'(1)), \ldots, (g'_i, \Gamma'(i)), (g'_{i+1}, \Gamma(i+1)), \ldots, (g'_n, \Gamma(n))). \tag{22e}$$

5 Numerical Results

Numerical simulations of our proposed model have been done for the parameters specified in Table 1. The most important results are displayed in Figure 5, where a sequential emergence of modularity is easily observed. In particular, it is postulated that a subset P_A is created from the first five characters, whereas a subset P_B is created from the remaining characters. The initial population was created randomly, restricting the maximal size of Γ-subsets to 5. Random epigenotypes have 8–12 crosslink mappings. This relatively large number of crosslink mappings decreased in the course of evolution to its minimal value 2. Plots of effective fitness and numbers of crosslinks in temporarily best solutions are displayed in Figures 6 and 7. Figure 8 shows a histogram

Table 1. Numerical values of parameters used in numerical simulations

No.	Parameter	Numerical value
1	n, number of genes	10
2	m, number of characters	10
3	m_A, number of characters in module A	5
4	$\lvert U(x)\rvert$, cardinality of neighborhood	100
5	Probability of one-gene mutation in calculation of effective fitness	0.1
6	$\lvert P\rvert$, size of population	100
7	Probability of one-mapping mutation in GA	0.01
8	Probability of one-point crossover in GA	0.5

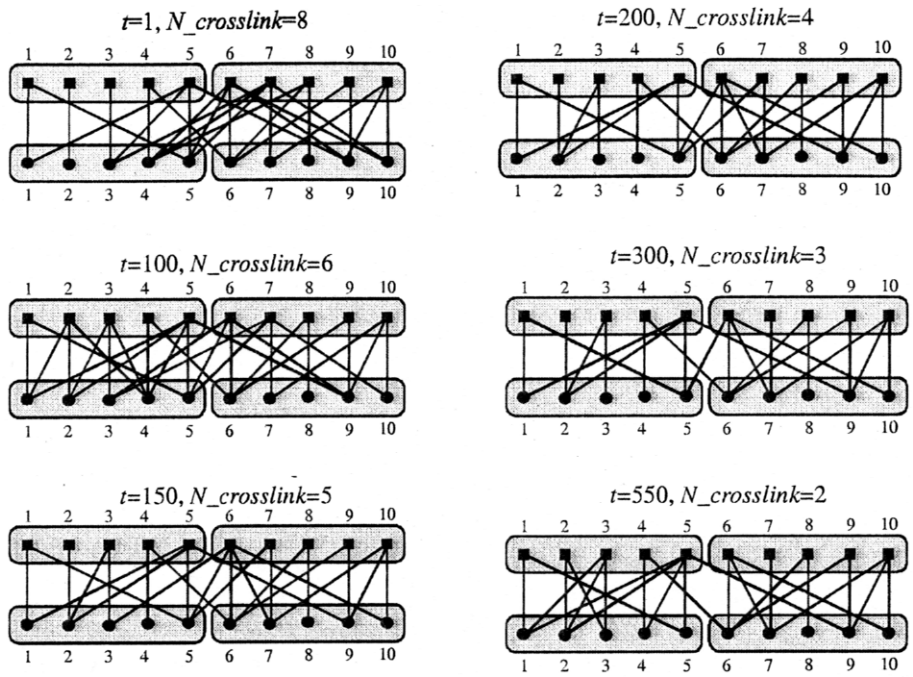

Figure 5. A sequence of six best solutions as they appeared during one run of the evolution process. The best solution at the beginning of evolution was characterized by eight crosslinks. This number, which characterizes the modularity of our model system, decreased monotonically through the evolution; its minimal value, 2, was achieved after 470 evolutionary epochs.

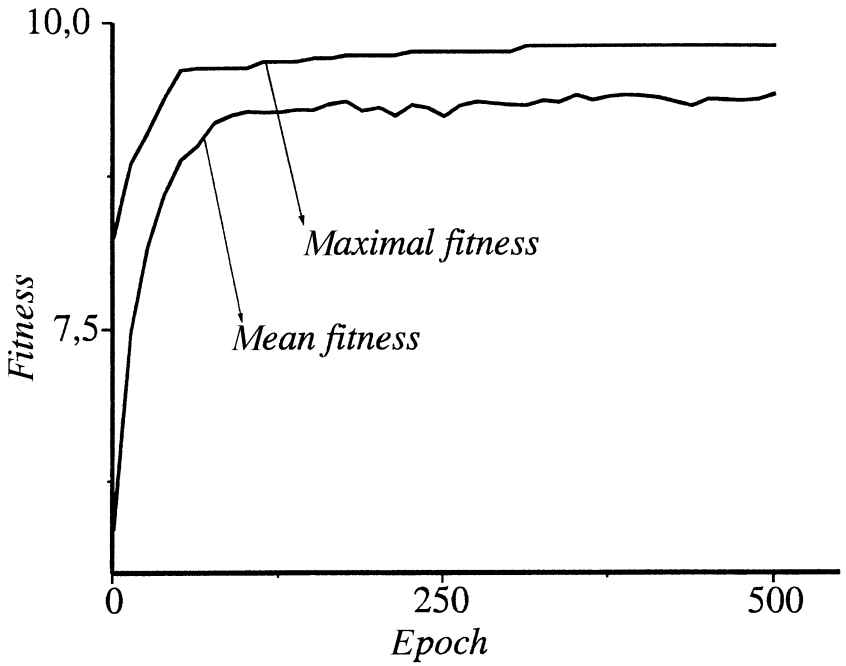

Figure 6. Plots of mean and maximal fitness through a whole Darwinian evolution. Both mean and maximal fitness increase almost monotonically to their maximal values.

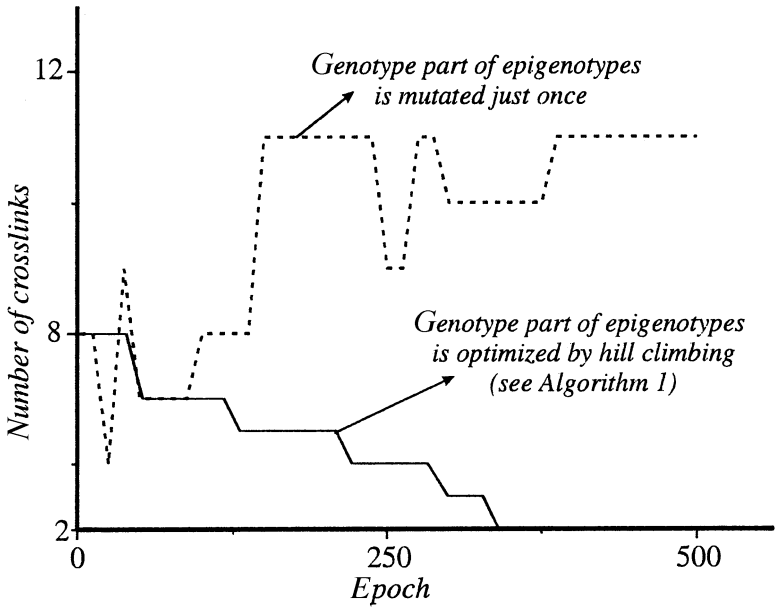

Figure 7. Plot of number of crosslinks (solid line) of the temporarily best solution in the course of Darwinian evolution. This number decreases monotonically from its maximal value 8 to its minimal value 2. For illustration, this plot is accompanied by a plot (dashed line) for Darwinian evolution when only one random mutation of the genotype part of epigenotypes was performed in stage 1. In this case we see that the number of crosslinks fluctuates in the course of evolution to its maximal value 11; consequently, modularity does not emerge.

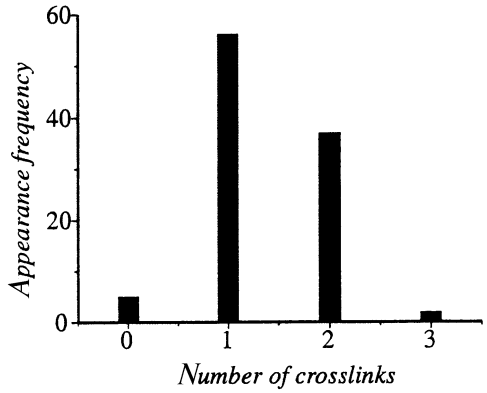

Figure 8. Histogram of 100 trial calculations. The best results were most frequently composed of epigenotypes containing one crosslink.

resulting from 100 trial calculations that were initiated by different initial populations of epigenotypes. We see that the most frequent result was an epigenotype that contained one crosslink.

6 Discussion

In this article we have described a novel method that allows us to study the emergence of modularity in the course of Darwinian evolution. The basic feature of this method is the application of composite epigenotypes with two parts: The first one is a binary

genotype, whereas the second one corresponds to a mapping of genes onto phenotype characters. For such generalized epigenotypes the modularity is determined in the following intuitive way: The genes are divided into two subgroups. Simultaneously with this decomposition there is specified also an accompanying decomposition of phenotype characters. We expect that for epigenotypes with modular structures the genes from one group will be mapped onto characters from the respective group, and crosslink mappings will be maximally suppressed. A fundamental question for the whole of evolutionary biology (and also for evolutionary algorithms and connectionist cognitive sciences) is: "What is the mechanism of evolutionary emergence of epigenotypes with modular structures?" If we accept the philosophy of Darwinian evolution, we may ask whether there exists a selection pressure that gives rise to direct emergence of modularity. The answer is definitely no [26]; the evolutionary algorithms are unable to produce modularity directly. The solution presented in this article, which overcomes this negative result, is to separate the variation of the modularity (i.e., mutations and optimization of pleiotropy) from the variations of the genotype part, and give different emphasis to these processes.

The difference in time scales for these variations offers an evolutionary tool that is potentially able to produce the emergence of modularity. On the other hand, simple evolutionary algorithms are unable to produce epigenotype modularity. This is the main reason why we have turned our attention to more sophisticated versions of evolutionary algorithms endowed with the concept of short time scale genotype evolution and long time scale genotype–phenotype-map optimization.

In the present paper we have considered congruence between modularity and directly selectable properties (after [26]) as the unavoidable prerequisite for the emergence of modularity. However, when a complex problem is being solved, this prerequisite alone may not ensure the emergence of modularity. Problems of greater complexity, such as the evolution of species, may be too much even for a standard evolutionary algorithm, which might be otherwise considered as the best algorithm for the optimization of complex problems with unknown structure.

The hypothesis supported in this article is that one possible basic cause of the emergence of modularity in a complex problem (apart from the congruence with directly selectable properties [26] mentioned above) is a two-speed mechanism of evolution. One mechanism works infrequently at the phenotype mapping level, and the other, more frequent mechanism is optimized at the genotype level. It has to be stressed that the fitness function did not include the modularity level explicitly; it was included only implicitly in that the optimization problem was separable into two suboptimization problems. The application of two time scales in optimization strategy, which achieved a fair level of modularity of the solution, was compared with a single-time-scale optimization strategy (where mutations in genotypes and pleiotropies occurred on the same order), which failed to achieve modularity. At the same time the single-time-scale optimization yielded a very bad value of the optimized function, compared with the two-level optimization.

The two-time-scale optimization used in our approach is similar to the so-called memetic algorithms [22]. However, the variations in the memetic algorithms were not used for different parts of the optimized structure, and these algorithms were neither designed nor used with an aim to achieve even partial biological plausibility.

The biological plausibility of secondary selection in genetic variations is supported by other work [5, 12–14, 20, 21].

We may say that the emergence of modularity of our epigenotype is sustained by suppression of the interference between different modules. The fitness is additive with respect to modules.

Acknowledgments

This work was supported by grants 1/7336/20 and 1/8107/01 of the Scientific Grant Agency of the Slovak Republic. The authors would like to thank both anonymous referees for their important notes leading to improvement of the manuscript.

References

1. Altenberg, L. (1994). Evolving better representations through selective genome growth. In *Proceedings of the IEEE World Congress on Computational Intelligence* (pp. 182–187).
2. Altenberg, L. (1994). The evolution of evolvability in genetic programming. In K. Kinnear (Ed.), *Advances in genetic programming* (pp. 47–74). Cambridge, MA: The MIT Press.
3. Altenberg, L. (1995). Genome growth and the evolution of the genotype-phenotype map. In W. Banzhaf & F. H. Eeckman (Eds.), *Evolution and biocomputation: Computational models of evolution* (pp. 205–259). Berlin: Springer-Verlag.
4. Altenberg, L. (1997). B2.7.2. NK fitness landscapes. In T. Baeck, D. F. Fogel, & Z. Michalewicz (Eds.), *Handbook of evolutionary computation*. New York: Oxford University Press.
5. Altenberg, L., & Feldman, M. W. (1987). Selection, generalized transmission, and the evolution of modifier genes. I. The reduction principle. *Genetics, 117*, 559–572.
6. Ancel, L. W., & Fontana, W. (2000). Plasticity, evolvability and modularity in RNA. *Journal of Experimental Zoology (Molecular and Developmental Evolution), 288*, 242–283.
7. Ancel, L. W., & Fontana, W. (in press). Evolutionary lock-in and the origin of modularity in RNA structure. In [11].
8. Bonner, J. T. (1988). *The evolution of complexity*. Princeton, NJ: Princeton University Press.
9. Calabretta, R., Nolfi, S., Parisi, D., & Wagner, G. P. (1998). A case study of the evolution of modularity: Towards a bridge between evolutionary biology, artificial life, neuro- and cognitive science. In C. Adami, R. Belew, H. Kitano, & C. Taylor (Eds.), *Artificial Life VI*. Cambridge, MA: MIT Press.
10. Calabretta, R., Nolfi, S., Parisi, D., & Wagner, G. P. (2000). Duplication of modules facilitates the evolution of functional specialisation. *Artificial Life, 6*, 69–84.
11. Callebout, W., & Raskin-Gutman, D. (Eds.) (in press). *Modularity: Understanding the development and evolution of complex natural systems*. Cambridge, MA: MIT Press.
12. Eshel, I. (1973). Clone-selection and optimal rates of mutation. *Journal of Applied Probability, 10*, 728–738.
13. Eshel, I. (1973). Clone-selection and the evolution of modifying features. *Theoretical Population Biology, 4*, 196–208.
14. Feldman, M. W., & Liberman, U. (1986). An evolutionary reduction principle for genetic modifiers. *Proceedings of the National Academy of Sciences of the U.S.A., 83*, 4824–4827.
15. Fonesca, A. M., & Fleming, P. J. (1997). C4.5 Multiobjective optimization. In T. Baeck, D. F. Fogel, & Z. Michalewicz (Eds.), *Handbook of evolutionary computation*. New York: Oxford University Press.
16. Goldberg, D. (1989). *Genetic algorthms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.
17. Holland J. (1992). *Adaptation in natural and artificial systems*. Cambridge, MA: MIT Press.
18. Kauffman, S. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford, UK: Oxford University Press.
19. Kernighan, S. W., & Lin, S. (1970). An efficient heuristic procedure for partitioning graphs. *Bell Systems Technical Journal, 2*, 291–307.
20. Maynard Smith, J. (1974). Recombination and the rate of evolution. *Genetics, 78*, 299–304.

21. Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research, 23*, 23–35.

22. Moscato, P. (1999). Memetic algorithms: A short introduction. In D. Corne, M. Dorigo, & F. Glover (Eds.), *New Ideas in Optimization* (pp. 219–234). London: McGraw-Hill.

23. Wagner, G. P. (1997). Homologues, natural kinds and the evolution of modularity. *American Zoologist, 36*, 36–43.

24. Wagner, G. P. (1995). Adaptation and the modular design of organisms. In F. Moran, A. Moran, J. J. Merelo, & P. Chacon (Eds.), *Advances in artificial life* (pp. 317–328). Berlin: Springer-Verlag.

25. Wagner, G. P., & Altenberg, L. (1997). Complex adaptations and the evolution of evolvability. *Evolution, 51*, 967–976.

26. Wagner, G. P., Mezey, J., & Calabretta, R. (in press). Natural selection and the origin of modules. In [11].

27. Wright, S. (1931). Evolution in Mendelian populations. *Genetics, 16*, 7–159.

Appendix: Details of Kauffman's NK Function Used in Our Simulations

Kauffman's function has become in recent models a standard testbed, an example of a function that is hard to optimize. Therefore we do not go into its statistical details. Its main characteristic is that it is *rugged*, that is, a small change in the argument causes a large change in the evaluation.

The functions φ in Equations 13 and 14b are defined formally so that they depend on three sets of arguments: both upper and lower indices, and binary arguments represented as nonnegative integers

$$c_j^i(k) = \varphi_j^{(i)}(\underbrace{g_{l_1}, g_{l_2}, \ldots, g_{l_j}}_{k}) \tag{23a}$$

where $k = \text{integer}(g_{l_1}, g_{l_2}, \ldots, g_{l_j})$. The function integer ( ) transforms the binary string $(g_{l_1}, g_{l_2}, \ldots, g_{l_j})$ into a decimal integer. Theoretically, coefficients $c_j^i(k)$ may be randomly generated and then stored in the form of a lookup table, but its dimensions may be very large. Therefore we turn our attention to another possibility, which overcomes the dimensionality problem: the coefficients $c_j^i(k)$ are determined pseudorandomly [4] by

$$c_j^i(k) = \text{rand}(\underbrace{1000i + 100j + k}_{l}), \tag{23b}$$

where rand(I) is a pseudorandom uniform generator of real numbers from an open interval (0,1); the integer I is its deterministic initiator (*RandSeed*) such that

$$I \neq I' \quad \Rightarrow \quad \text{rand}(I) \neq \text{rand}(I'). \tag{24}$$

This requirement severely restricts the standard random number generators to be used; therefore we have used in our calculations the most robust algorithms (with a long period).

For a better understanding of this complicated evaluation of epigenotypes by fitness we use the mapping outlined in Figure 3. We get

$$\begin{aligned} f(x) &= \varphi_4^{(1)}(\underbrace{1010}_{10}) + \varphi_3^{(2)}(\underbrace{011}_{3}) + \varphi_1^{(3)}(1) + \varphi_1^{(4)}(1) + \varphi_1^{(5)}(0) \\ &= c_4^1(10) + c_3^2(3) + c_1^3(1) + c_1^4(1) + c_1^5(0) \\ &= \mathrm{rand}(1410) + \mathrm{rand}(2303) + \mathrm{rand}(3101) + \mathrm{rand}(4101) + \mathrm{rand}(5100). \end{aligned} \tag{25}$$